**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K | A2

(11) International Publication Number: **WO 99/48909**

(43) International Publication Date: 30 September 1999 (30.09.99)

(21) International Application Number: PCT/US99/02142

(22) International Filing Date: 1 February 1999 (01.02.99)

(30) Priority Data:
60/073,223 30 January 1998 (30.01.98) US

(71)(72) Applicants and Inventors: GREISMAN, Harvey, A. [US/US]; Cambridge, MA (US). PABO, Carl, O. [US/US]; Newton, MA (US).

(74) Agents: GARRETT, Arthur, S. et al.; Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P., 1300 I Street, Washington, DC 20005–3315 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*Without international search report and to be republished upon receipt of that report.*

(54) Title: A GENERAL STRATEGY FOR SELECTING HIGH–AFFINITY ZINC FINGER PROTEINS FOR DIVERSE DNA TARGET SITES

(57) Abstract

A method of creating new zinc finger proteins directed to specific DNA binding sites comprising adding one finger to the protein at a time.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites

## BACKGROUND OF THE INVENTION

Design of DNA-binding proteins that will recognize desired sites on double-stranded DNA has been a challenging problem. Although a number of DNA-binding motifs have yielded variants with altered specificities, zinc finger proteins related to TFIIIA (1) and Zif268 (2) appear to provide the most versatile framework for design. Modeling, sequence comparisons, and phage display have been used to alter the specificity of an individual zinc finger within a multifinger protein (3-7), and fingers also have been "mixed and matched" to construct new DNA-binding proteins (8, 9). These design and selection studies have assumed that each finger [with its corresponding 3-base pair (bp) subsite] can be treated as an independent unit (Fig. 1B). This assumption has provided a useful starting point for design studies, but crystallographic studies of zinc finger-DNA complexes (10-13) reveal many examples of contacts that couple neighboring fingers and subsites, and it is evident that context-dependent interactions are important for zinc finger-DNA recognition (3, 7, 8). Existing strategies have not taken these interactions into account in the design of multifinger proteins, and this may explain why there has been no effective, general method for designing high-affinity proteins for desired target sites.

"Mix and match" design strategies have, so far, been limited to binding sites in which the primary strand (Fig. 1B) contains at least one guanine within each 3-bp subsite (3, 8, 9). The

affinities of designed zinc finger proteins also have varied widely, and some $K_d$'s have been in the micromolar range (8, 9). Subtle, context-dependent interactions (which provided the motivation for our protocol) may have a critical cumulative effect when optimizing multifinger proteins: A modest (10-fold) increase in affinity for each finger may yield a substantial (1000-fold) increase in affinity for a three-finger protein.

## DETAILED DESCRIPTION

We have developed a selection strategy that can accommodate many of the context-dependent interactions between neighboring fingers and subsites. Our strategy involves gradual assembly of a new zinc finger protein at the desired binding site--adding and optimizing one finger at a time as we proceed across the target site. In one embodiment, we use the Zif268 structure (10, 13) as our framework and randomize six potential base-contacting positions in each finger (Fig. 1, A and D).

Our protocol includes three selection steps (Fig. 2), one of each finger of the new protein: (i) A finger that recognizes the 3' end of the target site is selected by phage display (Fig. 2A). Examples of the technique of phage display have been published at e.g. U.S. Patent No. 5,223,409 issued June 29, 1993, U.S. Patent No 5,403,484 issued April 4, 1995, and 5,571,698 issued November 5, 1996, incorporated herein by reference. At this stage, two wild-type Zif fingers are used as temporary anchors to position the library of randomized fingers over the target site, and we use a hybrid DNA site that has Zif subsites fused to the target site. (ii) The selected finger is retained as part of a "growing" protein and, after the distal Zif finger is discarded, phage display is used to select a new finger that recognizes the central region of the

target site (Fig. 2B). (iii) Finally, the remaining Zif finger is discarded, and phage display is used to select a third finger that recognizes the 5' region of the target site (Fig. 2C). Optimization of this finger yields the new zinc finger protein.

Figure 1A depicts the amino acid sequence and secondary structure of the Zif268 zinc fingers. [Adapted from (10)] Randomized positions (circled) correspond to residues -- 1, 1, 2, 3, 5, and 6 in each of the α helices and include every position that makes a base contact in one of the known zinc finger-DNA complexes (10-13). The wild-type Zif268 sequence was retained at all other positions in the new proteins. Key base contacts (solid arrows) in the Zif268-DNA complex are depicted in Figure 1B (10, 13). Most of the bases contacted are located on the primary (guanine-rich) strand (boldface). Each finger makes several base contacts with its 3-bp subsite (dashed boxes), but also makes important base and phosphate contacts in flanking subsites. The 1.6 Å structure (13) shows that the aspartic acid at position 2 in finger 2 contacts a cytosine that is just outside the canonical 3-bp subsite. Analogous contacts from position 2 in the other fingers (dashed arrows) have less favorable hydrogen-bonding geometry, but binding site selections (19) suggest that these contacts may contribute to recognition. Contacts made by Tramtrack (11) and GLI (12) also include bases and phosphates outside the canonical 3-bp subsites. Figure 1C depicts DNA sequences of the sites used in our selections. The TATA box is from the adenovirus major late promoter (20), the p53 binding site is from the human $p21^{WAF1/CIP1}$ promoter (18), and the NRE is from the human apolipoprotein AI promoter (21). One strand of each duplex site is shown. The structure of the wild-type Zif268 zinc finger–DNA complex is depicted in Figure 1D (10, 13). The DNA is gray, and a ribbon trace of the three zinc fingers is shown in red (finger 1), yellow (finger 2), and purple (finger 3). The 18 residues that

were randomized in this study (van der Walls surfaces shown in blue) occupy the major groove of the DNA and span the entire length of the binding site. [Image created with Insight II (Biosym Technologies, San Diego, California)]

Our strategy ensures that the new fingers are always selected in a relevant structural context. Because of an intact binding site is present at every stage, and because our selections are performed in the context of a growing protein-DNA complex, our method readily optimizes context-dependent interactions between neighboring fingers and subsites and naturally selects for fingers that will function well together. To ensure that the selected proteins will bind tightly and specifically to the desired target sites, we performed all selections in the presents of calf thymus competitor DNA (3 mg/ml) (14). This serves to counterselect against any proteins that bind promiscuously or prefer alternative sites, and our protocol thus directly selects for affinity as well as specificity of binding. Assuming that the calf thymus DNA has one potential binding site per base (that is, binding could conceivably occur in any register on either strand) a 3 mg/ml solution of DNA corresponds to a 0.01 M solution of potential binding sites. (Our specific site is present at 40 nM). If the DNA sequence of this competitor were random, each of the $4^9$ (= 262.144) possible 9-bp sites would be present, with an average concentration of about 40 nM.

## EXAMPLE 1

An overview of a protocol that successively selects finger 1, finger 2, and finger 3 to create a new zinc finger protein is depicted in Figure 2. Fingers that are present in the phage libraries used in these steps are indicated on the left side of each panel.

Each cassette encodes one of the Zif268 fingers (Fig. 1A), and randomized codons have A/C/G at the first position, A/C/G/T at the second position, and C/G at the third position. These randomized codons allow 16 side chains at each position (all residues except Cys, Phe, Tyr, and Trp) and they do not give any termination codons. Each cassette encodes a maximum of $16^6$ (= $1.7 \times 10^7$) different zinc finger sequences represented by $24^6$ (= $1.9 \times 10^8$) different DNA sequences. All phage display libraries contained between $5.6 \times 10^8$ and $1.9 \times 10^9$ clones. After the finger 1 selections (Fig. 2A), double-stranded DNA was purified from $\geq 10^5$ optimized phagemids, and the first wild-type Zif finger was removed; transformed colonies ($\geq 10^7$) were pooled, and purified DNA from this pool was used to remove the remaining wild-type finger from the selected pool and to construct the finger 3 library. To accommodate the restriction sites used in these cloning steps (17), we changed residues in the COOH-terminal linker of each randomized finger to TGESR for one round of selections; wild-type residues were restored when the next cassette was added.

In Figure 2, "Zif1" and "Zif2" indicate wild-type Zif268 fingers. R indicates a randomized finger library, and asterisk indicates a selected finger. Small horizontal arrows indicate the multiple cycles of selection and amplification used when selecting each finger by phage display.

Phage display was performed in an anaerobic chamber to ensure proper folding of the zinc fingers (4, 14). Five to eight cycles of selection and amplification were performed for each finger, and retention efficiencies plateaued at values ranging from ~0.2 to 3% of input phage (14, 17). Binding reactions for the p53 finger 3 selections contained the nonbiotinylated duplex competitor 5'-CCCTTGGAACATGTTCCTGATCGCGG-3' (17). [The p53 target site is

pseudosymmetric (Fig. 1C) (18), and we wanted to avoid inadvertently selecting a zinc finger protein that would bind to the opposite strand]. The biotinylated sites used in the TATA box selections are shown in Fig. 2, and the sites used for the other selections (17) were designed in a similar manner; we altered the Zif268 subsites when they were no longer needed (Fig. 2 B and C) and removed any cryptic binding sites that resembled the binding site of interest.

The right side of each panel shows the binding sites used in selections with the TATA site and indicates the overall binding mode for the selected fingers [each DNA duplex has biotin (not shown) attached at the 3' end of the upper strand]. Vertical arrows indicate how fingers selected in earlier steps are incorporated into the phage libraries used in later steps and reselected to optimize affinity and specificity in the new context. Our protocol actually was designed so that a sublibrary of successful zinc finger sequences could be carried over from one selection step (Fig. 2, A or B) to the next. Preliminary sequencing data to analyze the "evolutionary history" of our selections (17) indicates that a set of finger 1 sequences was carried over into the step in Fig. 2B and that this step then selects for combinations of fingers that work well together.

A randomized finger 1 library was cloned into the pZif12 phagemid display vector. The pZif12 phagemid display vector (14) encodes a fusion protein that contains (i) Zif268 fingers 1 and 2 [residues 327 to 391 of the intact protein (2)]; (ii) a linker that introduces an amber codon; and (iii) residues 23 to 424 of the M13 gene III protein. The zinc finger region contains a set of restriction sites that were designed to facilitate the multiple cloning steps in our protocol (17). Selections with the library were performed in parallel at the TATA, p53, and NRE sites (14) (Fig. 2A). The wild-type Zif1 finger was removed, and a randomized finger 2 cassette was ligated to the appropriate vector pool and optimized by phage display (17). (Fig. 2B) The

remaining wild-type finger was removed, and a randomized finger 3 cassette was added and optimized by phage display (Fig. 2C). To construct the sites used in these selections, we fused the target strand with the higher purine content to the guanine-rich strand of the Zif268 site. Because of the overlapping base contacts that can occur at the junction of neighboring subsites (Fig. 1B), the 3' end of the target site (Fig. 1C) was aligned so that it overlapped with the Zif2 subsite.

## EXAMPLE 2

We tested our protocol by performing selections with a TATA box, a p53 binding site, and a nuclear receptor element (NRE) (Fig. 1C). These important regulatory sites were chosen because they normally are recognized by other families of DNA-binding proteins and because these sites are quite different from the guanine-rich Zif268 site and from sites that have been successfully targeted in previous design studies (14). After the multiple rounds of selections (Fig. 2) were completed, the final phage pools bound tightly to their respective target sites. DNA sequencing of eight clones from each pool revealed marked patterns of conserved residues (Fig. 3), and many of the selected residues (Arg, Asn, Gln, His, and Lys) could readily contribute to base recognition. Each set of proteins exhibits a clear gradient of sequence diversity across the three fingers (Fig. 3), but the finger 1 and finger 2 sequences were more diverse at intermediate stages of the optimization protocol (14). For example, after the first step (Fig. 2A), many of the TATA clones had Asn residues at position -1 or position 6 or in both locations. After the selections indicated in Fig. 2B, most clones had Gln at position -1 and Thr at position 6 of finger 1, and these residues also are present in a homologous natural finger that recognizes the same

subsite. Based on the Zif268 (Fig. 1B) and Tramtrack (11) structures, our alignments assume that residues at position -1 can contact the 3' base on the primary strand of the subsite, residues at position 3 can contact the central base, and residues at position 6 can contact the 5' base. Guanine bases in our sites appear to prefer Arg at positions -1 and 6, but His or Lys at position 3. Adenine bases appear to prefer Asn at position 3, but prefer Gln at position -1 and, to some extent, at position 6. Several of the subsites recognized by our optimized fingers (Fig. 3) also happen to appear in binding sites for the Tramtrack (11) and Gfi-1 zinc finger proteins [P.A. Zweidler-McKay, H.L. Grimes, M.M. Flubacher, P.N. Tschlis, *Mol. Cell. Biol.* 16, 4024 (1996), incorporated herein by reference], and we find remarkable similarities in the amino acid sequences of the corresponding recognition helices. These homologies include, but are not limited to, the canonical base-contacting residues at positions -1, 3, and 6. For example, finger 4 of the Gfi-1 protein and finger 1 of our NRE proteins appear to recognize the subsite 3'-ACT-5', and the Gfi-1 residues at positions -1, 1, 2, 3, 5, and 6 are QKSDKK (underlined residues match the consensus in the selected fingers). Finger 5 of Gfi-1 and finger 1 of the TATA proteins appear to recognize the subsite 3'-AAA-5', and the corresponding Gfi-1 residues are QSSNIT. (Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.)

## EXAMPLE 3

Because of the marked sequence conservation within each of the final phage pools, we used a single clone from each set for further analysis. The corresponding peptides were

overexpressed in *Escherichia coli* and purified. Zinc finger regions were subcloned in pET2d (Novagen), and the corresponding peptides (with end points as in Fig. 1A) were expressed in *E. coli* BL21 (DE3) and purified as described (4). Affinities of the peptides for their respective target sites were determined by electrophoretic mobility shift analysis.

Dissociation constants were determined essentially as described (4). However; (i) each $K_d$ was determined in the absence of competitor DNA; (ii) binding buffer contained 15 mM Hepes-NaOH (pH 7.9), 50 mM KCl, 50 mM potassium glutamate, 50 mM potassium acetate, 5mM $MgCl_2$, 20 μM $ZnSO_4$,acetylated bovine serum albumin (100 μg/ml), 5% (v/v) glycerol, and 0.1% (w/v) NP-40; (iii) binding reactions contained 2 or 4 pM of the labeled site and were equilbrated for 1 hour; (iv) $K_d$ values were calculated from the slopes of Scatchard plots and represent the average of three independent experiments (SD values were all <60%); and (v) mobility shift assays were performed with double-stranded oligonucleotides containing TTT overhangs at the 5' -AGGGGGGCTATAAAAGGGGGT-3' (TATA box), 5'-GCTGTTGGGACATGTTCGTGA-3' (p53 site), 5' -GCCGTCAAGGGTTCAGTGGGG-3' (NRE site), and '5 -CCAGTAGCGGGGGCGTCCTCG-3' (Zif268 site).

The measured dissociation constants ($K_d$'s) were 0.12 nM for the TATA box, 0.11 nM for the p53 binding site, and 0.038 nM for the NRE. These new complexes are almost as stable as the wild-type Zif268-DNA complex ($K_d$ of 0.010 nM under these buffer conditions).

Apparent $K_d$'s for nonspecific DNA were estimated by competition experiments with calf thymus DNA.

For competition experiments, 8 pM of labeled specific oligonucleotide was mixed with binding buffer containing successive twofold dilutions of calf thymus competitor DNA. An

equal volume of binding buffer that contained a fixed amount of protein (sufficient for a 50 to 80% mobility shift in the absence of competitor DNA) was added, after which the reaction mixtures were incubated for ≥ 1 hour and then subjected to gel electrophoresis (4). $K_d$ (in μg/ml) was calculated from the slope of a $C_t\theta$ versus plot, using the equation:

$$C_t\theta=[\frac{-K_d^{ns}}{1-\theta_o}]\theta+[\frac{K_d^{ns}}{(1-\theta_o)/\theta_o}]$$

where θ is the fraction of specific site bound by protein in the presence of competitor DNA (at concentration $C_t$), and $\theta_o$ is the fraction bound in the absence of competitor. This equation was derived from equation 3 of S.Y. Lin and A.D. Riggs (J. Mol. Biol. 72, 671 (1972), incorporated herein by reference]. Each $K_d^{ns}$ value represents the average of six plots (three plots in two independent experiments). All SD values were <25%. When calculating $K_d^{ns}/K_d$, we assumed that each base in the calf thymus DNA represents the beginning of a potential binding site. A simple estimate for the specificity of these new zinc finger proteins can be made by taking various powers of $4^n$ and comparing these numbers with the measured specificity ratios. All of our new proteins have specificity ratios that lie between $4^7$ (= 16.384) and $4^8$ (= 65.536). This indicates that our proteins--like Zif268 itself--can effectively specify 7 to 8 bp in the target DNA sites.

Ratios of the nonspecific to specific dissociation constants ($K_d^{ns}/K_d$) indicate that the peptides selected for the TATA box, p53 binding site, and NRE discriminate effectively against nonspecific DNA (preferring their specific sites by factors of 25,000, 54,000, and 36,000, respectively). These ratios are similar to the specificity ratio of 31,000 that we measured for wild-type Zif268. Taken together, the affinities and specificities of the new proteins indicate that they bind as well as many natural DNA-binding proteins.

## EXAMPLE 4

Figure 3 depicts amino acid sequences of new zinc finger proteins that recognize (A) the TATA box, (B) the p53 binding site, and (C) the NRE. Residues selected at each of the six randomized positions are shown. Four of the eight p53 clones had a conservative Ser – Thr mutation at position -2 in finger 2; in all other clones, residues outside the randomized regions were identical to those in wild-type Zif268. Six or more of the eight clones in each phage pool encode unique zinc finger proteins. A box indicates the clone that was overexpressed and used for binding studies. Residues that are fully conserved (eight of eight clones) are shown in boldface; residues that are partially conserved (four or more of eight) are denoted by lowercase letters in the consensus sequence below the set of clones. Modeling suggests that these new zinc finger proteins (including those that recognize the TATA box) can bind to B-form DNA. Each panel indicates how the fingers could dock with a canonical 3-bp spacing (dashed boxes), and dashed arrows indicate plausible base contacts. Recent data from studies of a designed zinc finger protein provide precedence for many of these contacts (22). Detailed modeling suggests many additional contacts (not shown), including some that couple neighboring fingers and

subsites. For the p53 site, there is an alternative, equally plausible, docking arrangement with a 4-bp spacing for one of the fingers. In the alternative arrangement, p53 finger 2 spans a 4-bp subsite (3'-ACAG-5') and finger 3 recognizes the adjacent 3'-GGT-5' subsite. A similar spacing occurs at one point in the GLI-DNA complex (12). A section of the NRE site shows a 5 of 6 bp match (undefined) with the Tramtrack binding site, and these matching segments happen to be aligned such that the new fingers bind in the same register as the Tramtrack fingers (11). Every Tramtrack residue that contacts one of the matching bases (solid arrows) was recovered in our selections. Two residues that do not directly contact the DNA in the Tramtrack complex were also recovered (at positions 5 and 6 in NRE finger 3).

Many discussions of zinc finger-DNA recognition have considered the idea of a "code" that specifies which positions along the α helix contact the DNA and which side chain-base interactions are most favorable at each position (5, 15). There are recurring patterns of contacts in some zinc finger proteins (10, 11), and similar patterns are apparent in the proteins we selected (Fig. 3). Thus, when adenine or guanine occurs in the primary strand of one of our binding sites (the strand corresponding to the guanine-rich strand of the Zif268 site), there often is a conserved residue at position -1, 3, or 6 of the α helix that could form hydrogen bonds with this base. Related patterns have been discussed in previous design and selection studies (3-6). There also are strong "homologies" between the zinc fingers we have selected and natural zinc fingers that may recognize the same subsites (Fig. 3).

Such simple patterns are not seen at other positions in our selected proteins. Thus, we found no simple patterns of residues at positions 1, 2, and 5 of the α helix, and when thymine or cytosine occurs on the primary strand (Fig. 3), we found no simple pattern of potential contacts

from residues at positions -- 1, 3, and 6. However, there still are numerous instances in which residues at these positions are highly conserved within a particular set of proteins (Fig. 3), and we infer that many of these considered residues make energetically significant contributions to folding or binding.

Given the remarkable homology with Tramtrack (Fig. 3), it seems plausible that the Ser and Asp residues at position 2 in NRE fingers 2 and 3 may make the same contacts that corresponding residues make in Tramtrack fingers 1 and 2 (11). We also anticipate that the Lys at position 1 in finger 1 of the TATA box proteins may make a phosphate contact analogous to the contact made by Tramtrack finger 2.

Because no readily predicted pattern of coded contacts is apparent, we surmise that residues at these positions may be involved in more subtle, context-dependent interactions. In short, there still is no general code that can be used to design optimal zinc finger proteins for any desired target sequence or that can predict the preferred binding site of every zinc finger protein. There are several examples of zinc fingers that have appropriate residues (Arg, His, Asn, or Gln) at positions -1, 3, and 6, but do not make the expected coded contacts with their 3-bp subsites. Examples include some natural fingers, such as finger 3 of GLI (12) and finger 2 of ADR1 (7), as well as synthetic fingers designed to recognize particular subsites (3). As noted by others (3, 7), context-dependent interactions may explain these effects.

Nonetheless, our sequential selection strategy should provide valuable information about potential patterns in zinc finger-DNA recognition, because it (i) makes few assumptions about the preferred spacing, docking, or contacts of the individual fingers; (ii) yields proteins with essentially wild-type affinities and specificities; (iii) yields sequences that match very will with

those of natural zinc finger proteins that recognize similar subsites; and (iv) can readily be adapted to pursue analogous studies with other TFIIIA-like zinc finger proteins.

The sequential selection strategy provides a general and effective method for design of new zinc finger proteins, and our success with a diverse set of target sites suggests that it should be possible to select zinc finger proteins for many important regulatory sequences. These proteins could then be fused with appropriate regulatory of effector domains for a variety of applications. The protocol also could be adapted to allow selection of proteins with four, five, or six fingers or to allow optimization of zinc fingers fused to other DNA-binding domains (16). Related selection methods might be developed for other families of multidomain proteins, including other DNA- and RNA-binding proteins, and possibly even modular domains involved in protein-protein recognition. The sequential selection strategy should open the field to a host of applications and studies, including tests to see how designer zinc finger proteins can be used in gene therapy.

## Background and Reference Articles

The articles listed are incorporated herein by reference:

1. J. Miller, A.D. McLachian, A. Klug, *EMBO J.* 4, 1609 (1985).

2. B.A. Christy, L.F. Lau, D. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* 85, 7857 (1988).

3. J. Nardelli, T. Gibson, P. Chamay, *Nucleic Acids Res.* 20, 4137 (1992); S.K. Thukral, M.L. Morrison, E.T. Young, *Mol. Cell. Biol.* 12, 2784 (1992); J.R. Desjarlais and J.M. Berg, *Proteins* 12, 101 (1992); *ibid.* 13, 272 (1992); *Proc. Natl. Acad. Sci. U.S.A.* 89, 7345 (1992); *ibid.* 91, 11099 (1994).

4. E.J. Rebar and C.O. Pabo, *Science* 263m 671 (1994).

5. Y. Choo and A. Klug, *Proc. Natl. Acad. Sci. U.S.A.* 91, 11163 (1994); *ibid.* p. 11168.

6. A.C. Jamieson, S.H. Kim, J.A. Wells, *Biochemistry* 33, 5689 (1994); H. Wu, W.P. Yang, C.F. Barbas, *Proc. Natl. Acad. Sci. U.S.A.* 92, 344 (1995).

7. W.E. Taylor et al., *Biochemistry* 34, 3222 (1995); C. Cheng and E.T. Young, *J. Mol. Biol.* 251, 1 (1995).

8. J.R. Desjarlais and J.M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* 90m 2256 (1993).

9. Y. Choo, I. Sanchez-Garcia, A. Klug, *Nature* 372, 642 (1994).

10. N.P. Pavletich and C.O. Pabo, *Science* 252, 809 (1991).

11. L. Fairall, J.W.R. Schwabe, L. Chapman, J.T. Finch, D. Rhodes, *Nature* 366, 483 (1993).

12. N.P. Pavletich and C.O. Pabo, *Science* 261, 1701 (1993).

13. M. Elrod-Erickson, M.A. Rould, L. Nekludova, C.O. Pabo, *Structure* 4, 1171 (1996).

14. E.J. Rebar, H.A. Greisman, C.O. Pabo, *Methods Enzymol.* 267, 129 (1996).

15. J.M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* 89, 11109 (1992); R.E. Kievit, *Science* 253, 1367 (1991); M. Suzuki, M. Gerstein, N. Yagi, *Nucleic Acids Res.* 22, 3397 (1994).

16. J.L. Pomerantz, P.A. Sharp, C.O. Pabo, *Science* 267, 93 (1995).

17. H.A. Greisman, thesis, Massachusetts Institute of Technology, Cambridge, MA (1997), MIT Libraries, Rm 14-0551, Cambridge, MA 02139-4307
Indexed in Dissertation Abstracts International Volume 58/04-B, p. 1692 (1997).

18. W.S. El-Deiry et al., *Cell* 15, 817 (1993); W.S. El-Deiry et al., *Cancer Res.* 55, 2910 (1995).

19. A.H. Swimoff and J. Milbrandt, *Mol. Cell. Biol.* 15, 2275 (1995).

20. E.B. Ziff and R.M. Evans, *Cell* 15, 1463 (1978).

21. J.A.A. Ladias and S.K. Karathanasis, *Science* 251, 561 (1991).

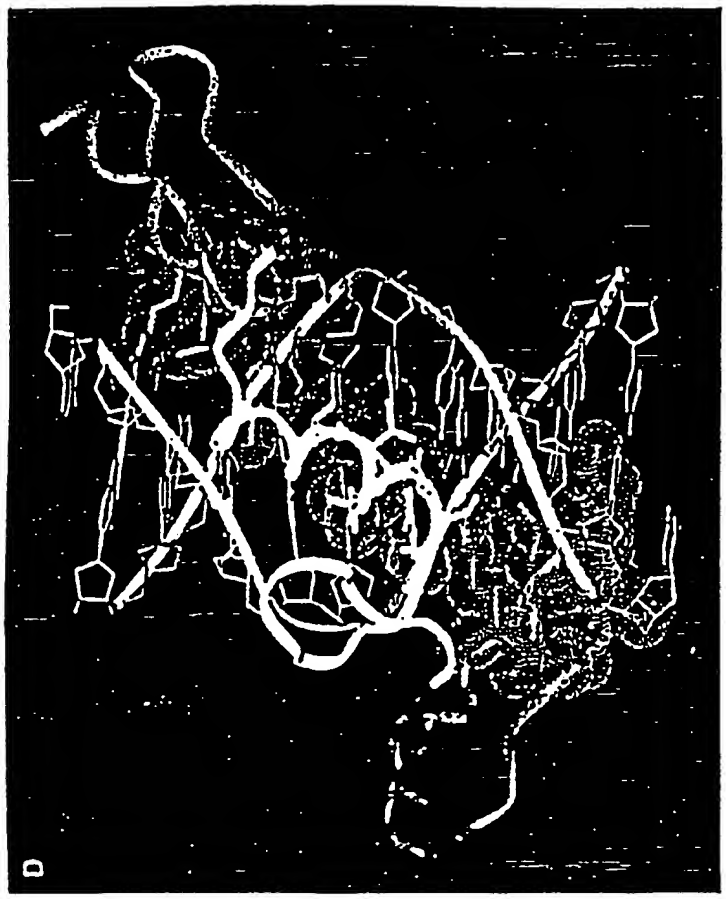
22. C.A. Kim and J.M. Berg, *Nature Struct. Biol.* 3, 940 (1996).

23. H.A. Creisman and C.O. Pabo, Science 275:657 (1997)

What is claimed is:

1. A method of creating new zinc finger proteins directed to specific DNA binding sites comprising adding one finger to the protein at a time.

Figure 1


D

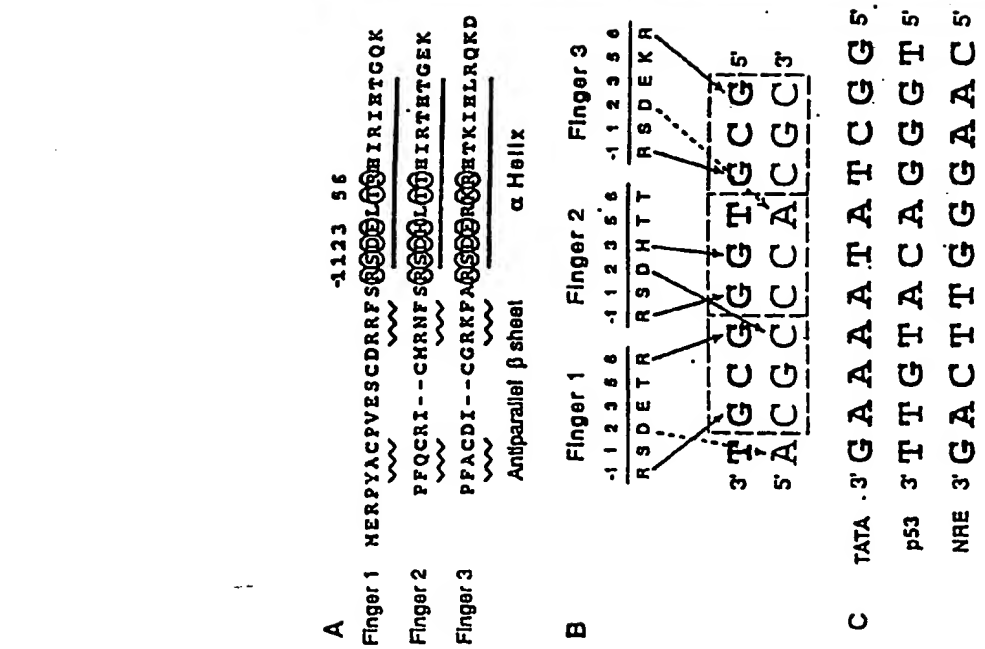

A
-1123 56
Finger 1 HERPYACPVESCDRRFSRSDELVRHHNMHIRHTGQK
Finger 2 PFQCRI--CMRNFSRSDNLKRHIRTHTGEK
Finger 3 PFACDI--CGRKFARSDERKRHTKIHLRQKD
Antiparallel β sheet
α Helix
B
Finger 1 Finger 2 Finger 3
-1 1 2 3 5 6 -1 1 2 3 5 6 -1 1 2 3 5 6
R S D E T R R S D H T T R S D E K R
3' T G C G G G T G C G 5'
5' A C G C C C A C G C 3'
C TATA 3' G A A A T A T C G G 5'
p53 3' T T G T A C A G G G T 5'
NRE 3' G A C T T G G G A A C 5'

Figure 2

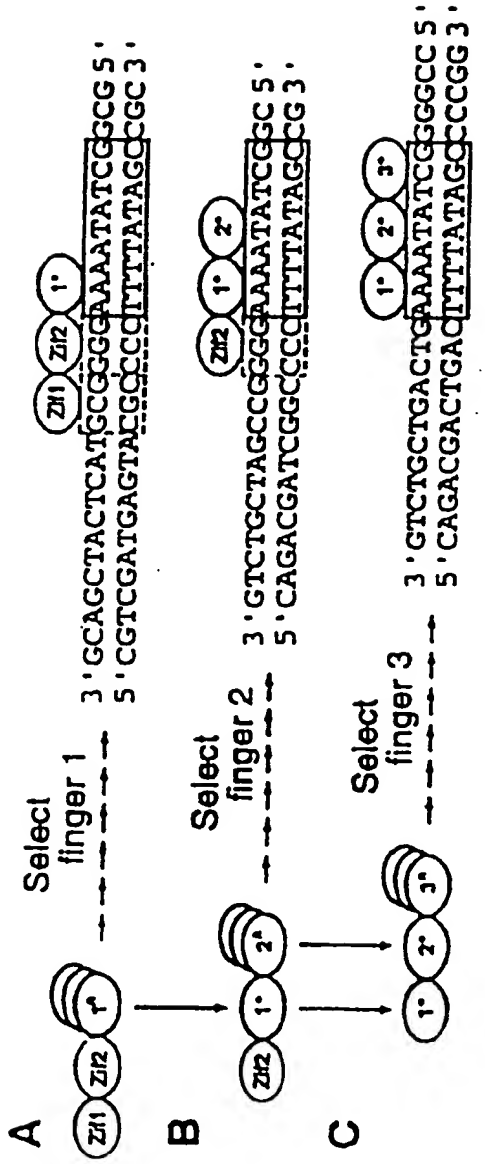
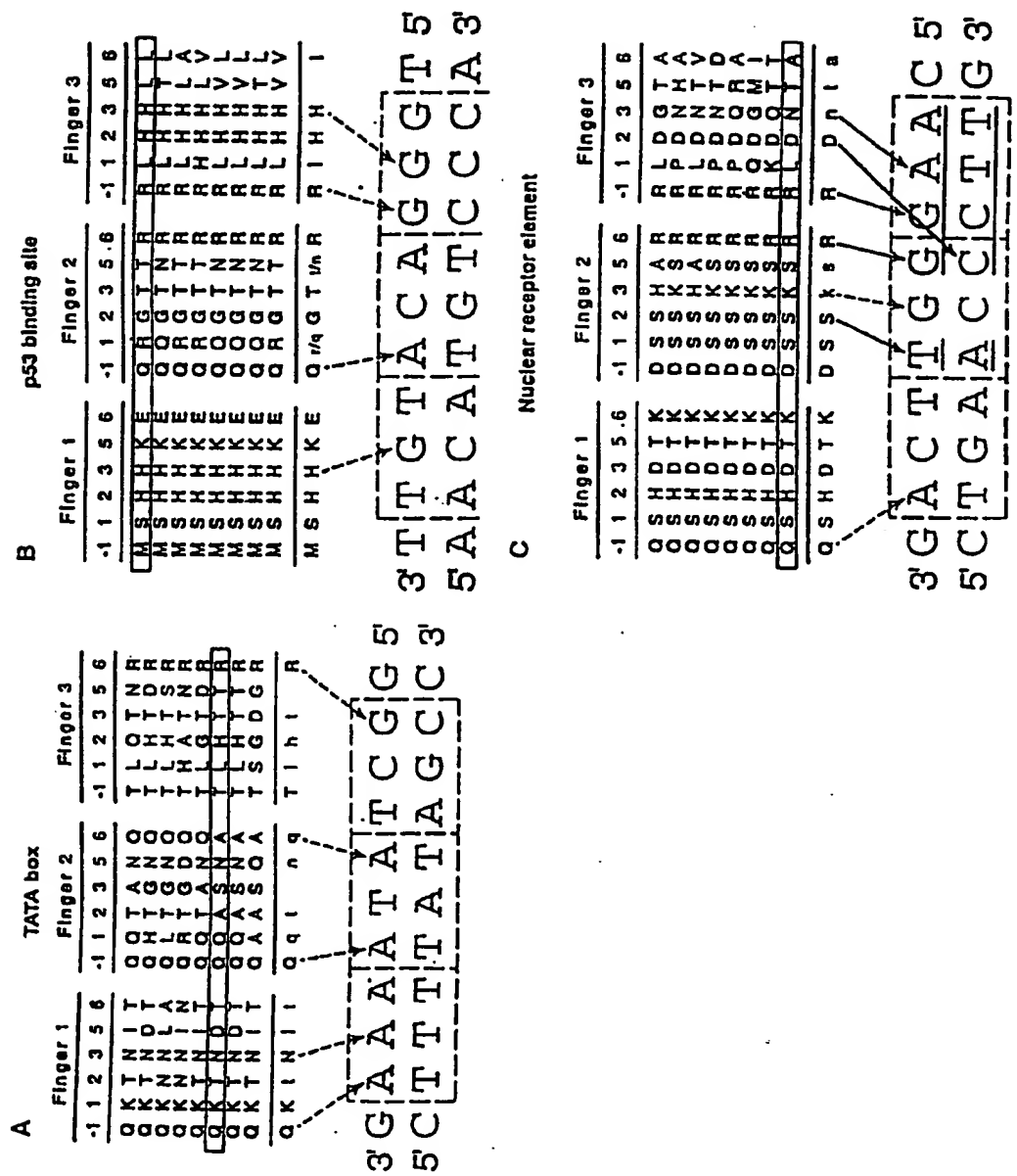

A
B
C
Select finger 1
Select finger 2
Select finger 3
Zf1
Zf2
1'
2'
3'
3'GCAGCTACTCATGCGGGGAAAATATCGGCG5'
5'CGTCGATGAGTACGCCCCTTTTATAGCCGC3'
3'GTCTGCTAGCCGGGGAAAATATCGGC5'
5'CAGACGATCGGCCCCTTTTATAGCCG3'
3'GTCTGCTGACTGAAAATATCGGGCC5'
5'CAGACGACTGACTTTTATAGCCCGG3'

Figure 3


A
TATA box
Finger 1
Finger 2
Finger 3
B
p53 binding site
Finger 1
Finger 2
Finger 3
C
Nuclear receptor element
Finger 1
Finger 2
Finger 3